

PATENT ABSTRACTS OF JAPAN

(11) Publication number : 11-242672
 (43) Date of publication of application : 07. 09. 1999

(51) Int. Cl. G06F 17/27
 G06F 17/21

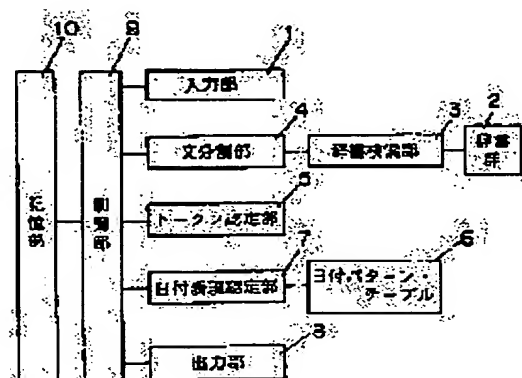
(21) Application number : 10-043234 (71) Applicant : MATSUSHITA ELECTRIC IND CO LTD
 (22) Date of filing : 25. 02. 1998 (72) Inventor : KINOSHITA HITOMI

(54) DEVICE AND METHOD FOR ANALYZING JAPANESE MORPHEME AND INFORMATION RECORDING MEDIUM

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a Japanese morpheme analyzer with which date expressions in a Japanese sentence can be collected as one token in morpheme analysis.

SOLUTION: This Japanese morpheme analyzer is provided with an input part 1 for inputting a prescribed Japanese character string, dictionary group 2 storing vocabulary information required for analyzing the morpheme of the Japanese character string, dictionary retrieving part 3 for retrieving the dictionary group 2 in which the Japanese character string is made as a key, sentence dividing part 4 for dividing the Japanese character string while using the dictionary retrieving part 3, token certifying part 5 for certifying individual unit divided at the sentence dividing part 4 as a token and a date expression certifying part 7 for certifying the date expression by collating the pattern of a date expression extracted in advance with the certified result at the token certifying part 5.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C) ; 1998, 2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11) 特許出願公開番号

特開平11-242672

(43) 公開日 平成11年(1999) 9月7日

(51) Int.Cl.⁶

識別記号

F I

G 0 6 F 17/27
17/21

G 0 6 F 15/20

5 5 0 E
5 9 0 J

審査請求 未請求 請求項の数 6 O L (全 8 頁)

(21) 出願番号 特願平10-43234

(22) 出願日 平成10年(1998) 2月25日

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 木下 ひとみ

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

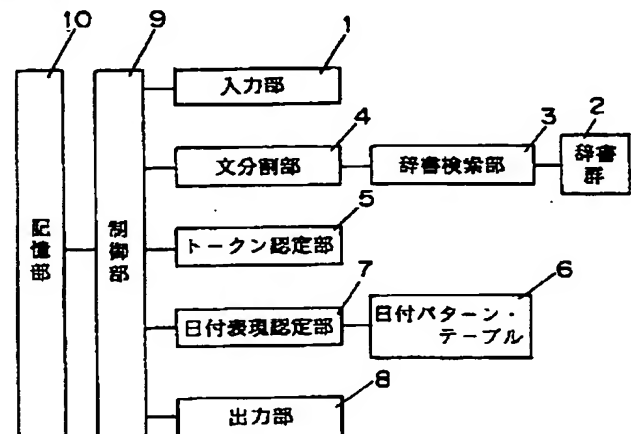
(74) 代理人 弁理士 滝本 智之 (外1名)

(54) 【発明の名称】 日本語形態素解析装置及び方法並びに情報記録媒体

(57) 【要約】

【課題】 形態素解析での日本語文における日付表現を1つのトークンとしてまとめることのできる日本語形態素解析装置を得る。

【解決手段】 所定の日本語文字列を入力する入力部1と、日本語文字列の形態素解析に必要な語彙情報が記憶された辞書群2と、日本語文字列をキーとして辞書群2を検索する辞書検索部3と、辞書検索部3を用いて日本語文字列を分割する文分割部4と、文分割部4において分割された個々の単位をトークンとして認定するトークン認定部5と、予め抽出された日付表現のパターンとトークン認定部5による認定結果とを照合して日付表現の認定を行う日付表現認定部7とを備えた日本語形態素解析装置とする。



【特許請求の範囲】

【請求項1】所定の日本語文字列を入力する入力手段と、
前記日本語文字列の形態素解析に必要な語彙情報が記憶された辞書群と、
前記日本語文字列をキーとして前記辞書群を検索する辞書検索手段と、
前記辞書検索手段を用いて前記日本語文字列を分割する文分割手段と、
前記文分割手段において分割された個々の単位をトークンとして認定するトークン認定手段と、
予め抽出した日付表現のパターンと前記トークン認定手段による認定結果とを照合して日付表現の認定を行う日付表現認定手段とを備えたことを特徴とする日本語形態素解析装置。

【請求項2】第1言語を日本語、第2言語を日本語以外の言語とした場合、前記第1言語の和暦を前記第2言語の西暦に、前記第1言語の月を前記第2言語の月に、前記第1言語の曜日を前記第2言語の曜日にそれぞれ変換して前記第2言語の日付表現文字列を生成する文字列変換手段を備えたことを特徴とする請求項1記載の日本語形態素解析装置。

【請求項3】日本語文字列を入力し、
前記日本語文字列の形態素解析を行ってこの日本語文字列を分割し、
分割された前記日本語文字列の個々の単位をトークンとして認定し、
予め抽出した日付表現のパターンとトークン認定結果とを照合して日付表現の認定を行うことを特徴とする日本語形態素解析方法。

【請求項4】第1言語を日本語、第2言語を日本語以外の言語とした場合、前記第1言語の和暦を前記第2言語の西暦に、前記第1言語の月を前記第2言語の月に、前記第1言語の曜日を前記第2言語の曜日にそれぞれ変換して前記第2言語の日付表現文字列を生成することを特徴とする請求項3記載の日本語形態素解析方法。

【請求項5】入力された日本語文字列の形態素解析を行うステップと、
形態素解析の結果を用いてこの日本語文字列を分割するステップと、
分割された前記日本語文字列の個々の単位をトークンとして認定するステップと、
予め抽出した日付表現のパターンとトークン認定結果とを照合して日付表現の認定を行うステップとが実行されることを特徴とする情報記録媒体。

【請求項6】第1言語を日本語、第2言語を日本語以外の言語とした場合、前記第1言語の和暦を前記第2言語の西暦に、前記第1言語の月を前記第2言語の月に、前記第1言語の曜日を前記第2言語の曜日にそれぞれ変換して前記第2言語の日付表現文字列を生成するステップ

が実行されることを特徴とする請求項5記載の情報記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、文字列として入力した日本語文の形態情報を出力する日本語形態素解析装置及び方法並びに情報記録媒体に関するものである。

【0002】

【従来の技術】ワープロのかな漢字変換や機械翻訳など、日本語を処理する場合には、まず、形態素解析を行う必要がある。形態素解析では、普通、単語をキーとしてその語彙情報を記憶した辞書を検索しながら、文字列を意味を持つ最小の単位、つまり形態素（以下、「トークン」という。）に分割し（トークン分割）、個々のトークンに品詞、活用などの形態情報を付加して、解析木として出力する。

【0003】

【発明が解決しようとする課題】この時、なるべく曖昧性のない状態で出力したり、より少ないトークン数で出力した方が、構文解析など以降の処理にとって効率が良い場合がある。

【0004】例えば、「平成9年10月25日土曜日」という文字列を、

「平成」＋「9」＋「年」＋「10」＋「月」＋「25」＋「日」＋「土曜日」

と出力すると、構文解析で日付の認定を行わなければならない。しかし、

「平成9年10月25日土曜日」

と出力すると、構文解析で日付の認定を行う必要はなくなる。

【0005】そこで、本発明は、形態素解析での日本語文における日付表現を1つのトークンとしてまとめることのできる日本語形態素解析技術を提供することを目的とする。

【0006】

【課題を解決するための手段】この課題を解決するために、本発明の日本語形態素解析装置は、所定の日本語文字列を入力する入力手段と、日本語文字列の形態素解析に必要な語彙情報が記憶された辞書群と、日本語文字列をキーとして辞書群を検索する辞書検索手段と、辞書検索手段を用いて日本語文字列を分割する文分割手段と、文分割手段において分割された個々の単位をトークンとして認定するトークン認定手段と、予め抽出した日付表現のパターンとトークン認定手段による認定結果とを照合して日付表現の認定を行う日付表現認定手段とを備えたものである。

【0007】これにより、形態素解析での日本語文における日付表現を1つのトークンとしてまとめることが可能になる。

【0008】

10

20

30

40

50

【発明の実施の形態】本発明の請求項 1 に記載の発明は、所定の日本語文字列を入力する入力手段と、日本語文字列の形態素解析に必要な語彙情報が記憶された辞書群と、日本語文字列をキーとして辞書群を検索する辞書検索手段と、辞書検索手段を用いて日本語文字列を分割する文分割手段と、文分割手段において分割された個々の単位をトークンとして認定するトークン認定手段と、予め抽出した日付表現のパターンとトークン認定手段による認定結果とを照合して日付表現の認定を行う日付表現認定手段とを備えた日本語形態素解析装置であり、形態素解析での日本語文における日付表現を 1 つのトークンとしてまとめることが可能になるという作用を有する。

【0009】本発明の請求項 2 に記載の発明は、請求項 1 記載の発明において、第 1 言語を日本語、第 2 言語を日本語以外の言語とした場合、第 1 言語の和暦を第 2 言語の西暦に、第 1 言語の月を第 2 言語の月に、第 1 言語の曜日を第 2 言語の曜日にそれぞれ変換して第 2 言語の日付表現文字列を生成する文字列変換手段を備えた日本語形態素解析装置であり、第 1 言語の日付表現を第 2 言語の日付表現に変換することが可能になるという作用を有する。

【0010】本発明の請求項 3 に記載の発明は、日本語文字列を入力し、日本語文字列の形態素解析を行ってこの日本語文字列を分割し、分割された日本語文字列の個々の単位をトークンとして認定し、予め抽出した日付表現のパターンとトークン認定結果とを照合して日付表現の認定を行う日本語形態素解析方法であり、形態素解析での日本語文における日付表現を 1 つのトークンとしてまとめることが可能になるという作用を有する。

【0011】本発明の請求項 4 に記載の発明は、請求項 3 記載の発明において、第 1 言語を日本語、第 2 言語を日本語以外の言語とした場合、第 1 言語の和暦を第 2 言語の西暦に、第 1 言語の月を第 2 言語の月に、第 1 言語の曜日を第 2 言語の曜日にそれぞれ変換して第 2 言語の日付表現文字列を生成する日本語形態素解析方法であり、第 1 言語の日付表現を第 2 言語の日付表現に変換することが可能になるという作用を有する。

【0012】本発明の請求項 5 に記載の発明は、入力された日本語文字列の形態素解析を行うステップと、形態素解析の結果を用いてこの日本語文字列を分割するステップと、分割された日本語文字列の個々の単位をトークンとして認定するステップと、予め抽出した日付表現のパターンとトークン認定結果とを照合して日付表現の認定を行うステップとが実行される情報記録媒体であり、形態素解析での日本語文における日付表現を 1 つのトークンとしてまとめることが可能になるという作用を有する。

【0013】本発明の請求項 6 に記載の発明は、請求項 5 記載の発明において、第 1 言語を日本語、第 2 言語を

日本語以外の言語とした場合、第 1 言語の和暦を第 2 言語の西暦に、第 1 言語の月を第 2 言語の月に、第 1 言語の曜日を第 2 言語の曜日にそれぞれ変換して第 2 言語の日付表現文字列を生成するステップが実行される情報記録媒体であり、第 1 言語の日付表現を第 2 言語の日付表現に変換することが可能になるという作用を有する。

【0014】以下、本発明の実施の形態について、図 1 から図 3 を用いて説明する。なお、これらの図面において同一の部材には同一の符号を付しており、また、重複した説明は省略されている。

【0015】（実施の形態 1）図 1 は、本発明の日付表現認定に関する実施の形態 1 における日本語形態素解析装置を示す機能ブロック図、図 2 は図 1 の日本語形態素解析装置の回路ブロック図である。

【0016】図 1 に示すように、本実施の形態の日本語形態素解析装置は、入力部（入力手段）1、文分割部（文分割手段）4、トークン認定部（トークン認定手段）5、日付表現認定部（日付表現認定手段）7 及び、出力部 8 を有している。また、これらを制御する制御部 9 及び所定情報を記憶する記憶部 10 が設けられている。さらに、文分割部 4 には辞書検索部（辞書検索手段）3 及び辞書群 2 が、日付表現認定部 7 には日付パターン・テーブル 6 が、それぞれ設けられている。

【0017】ここで、入力部 1 は、ユーザにより日本語文が文字列として入力される。なお、入力される日本語文字列は、たとえば漢字仮名混じり文とする。辞書群 2 は、日本語文字列をキーとして、読み情報、品詞、活用の種類など、形態素解析に必要な語彙情報を記憶する。辞書検索部 3 は、日本語文字列をキーとして、辞書群 2 を検索する。文分割部 4 は、辞書検索部 3 の結果を用いて入力文を辞書登録単位に分割する。トークン認定部 5 は、文分割部 4 で分割した個々の単位をトークンとして認定する。ここでは、隣り合う辞書未登録語を字種によってまとめたり、数字をまとめたりする処理が行われる。日付パターン・テーブル 6 は、日本語における日付表現のパターンを予め抽出し、テーブル化したものである。日付表現認定部 7 は、トークン認定部 5 で認定したトークンの並びと日付パターン・テーブル 6 とを照合し、入力文に含まれる日付表現を認定する。出力部 8 は、トークン認定部 5 及び日付表現認定部 7 の結果を形態素解析木として出力する。制御部 9 は、入力部 1、文分割部 4、トークン認定部 5、日付表現認定部 7 及び出力部 8 を制御し、また、入力部 1 で入力された文の形態情報を出力する。記憶部 10 は、入力部 1 で入力された文、文分割部 4 における分割結果、トークン認定部 5 ならびに日付表現認定部 7 における認定結果を記憶する。

【0018】図 2 に示すように、日本語形態素解析装置は、中央演算処理装置（以下、「CPU」という。）によって動作制御されるキーボード（マウスを含む）21、陰極線管ディスプレイ（以下、「CRT」とい

う。)22、ランダムアクセスメモリ(以下、「RAM」という。)24、プログラム記憶領域であるリードオンリーメモリ(以下、「ROM」という。)25を有している。ここで、入力部1はキーボード21により、記憶部10はRAM24により実現されている。また、日付パターン・テーブル6はROM25に、辞書群2は、RAM24、ROM25、2次記憶装置のいずれかに記憶されている。そして、辞書検索部3、文分割部4、トークン認定部5、日付表現認定部7、出力部8、制御部9の動作は、CPU23がRAM24およびROM25とデータのやりとりを行いながらROM25に記憶されたプログラムを実行することにより実現されている。

【0019】なお、本実施の形態では、CPU23はROM25に記憶されたプログラムを実行する形態となっているが、プログラムはCD-ROMなどの記録媒体27に記録されたものであっても構わない。

【0020】このような形態とした場合は、記録媒体27に記録されているプログラムをディスクドライブ26から読み取り、RAM24に展開したものをCPU23が実行する形態となる。

【0021】このような形態とすることにより、本発明を汎用コンピュータで容易に実現することができる。

【0022】(実施例1)図4は実施の形態1に対応した実施例1における、入力文から日付表現を認定する形態素解析処理の流れを示すフローチャート、図5は実施例1における日付表現認定処理の流れを示すフローチャート、図6は実施例1における日付表現認定の処理に用いられる日付パターン・テーブルを示す説明図、図7は実施例1における照合結果格納領域を示す説明図である。

【0023】まず、1文単位の入力を行う(ステップ4a)。ここでは、

「彼は、1970年6月28日に生まれた。」
なる文が入力されたものとする。

【0024】これにより、1文単位の入力があったかがチェックされ(ステップ4b)、入力があれば次の処理へ移り、入力が無ければステップ4aへ戻り、入力待つ。

【0025】次に、辞書群(図1)を用いて、入力文を辞書登録単位に分割する(ステップ4c)。ここでは、以下のように17の文字列に分割される(分割単位を/で区切って示している)。

【0026】彼/は/、/1/9/7/0/年/6/月/2/8/日/に/生まれ/た/。

このように分割されたならば、隣り合う辞書未登録語を字種によってまとめたり、数字をまとめたりするトークン認定処理を行う(ステップ4d)。入力文には、辞書未登録語は含まれていなかったため、ここでは、数字をまとめて1つのトークンとする。その結果、以下のよう

に13のトークンが認定される(トークンを/で区切って示している)。

【0027】彼/は/、/1970/年/6/月/28/日/に/生まれ/た/。

そして、変数posに先頭のトークンを指すポインタをセットする(ステップ4e)。

【0028】次に、変数posから始まる日付表現認定の処理を行う(ステップ4f)。ここで用いる日付パターン・テーブルを図6に示す。

10 【0029】図6(a)は、日付認定処理における日付表現のパーツ「年」「月」「日」「曜日」の認定で用いるテーブルである。テーブル中に「」付きで表記しているものは、トークン文字列そのものを意味している。また、「」が付いていないものは、トークンの種類を示している。例えば、年パターン1の元号には、「明治」「大正」「昭和」「平成」といった種類の文字列が存在する。ここで、数は数字のトークンを意味している。従って、年パターン1は、「平成」+「9」+「年」なるトークン・パターンを表現していることになる。図6

20 (b)は、日付表現のパターンを表したものである。1行目のパターンは、「平成9年10月25日土曜日」のように、「年」「月」「日」「曜日」全てが記述されるパターンを、5行目は、「10月25日」のように、「月」「日」のみが記述されるパターンを示している。

【0030】ここで、日付表現認定処理を図5を用いて説明する。まず、posが指すトークンの文字列が「年」か否かチェックし(ステップ5a)、「年」であればステップ5bへ、「年」でなければステップ5eへ移る。

30 【0031】文字列が「年」であれば、年パターン・テーブル(図6(a))とトークン列との照合を行い(ステップ5b)、一致したか否かチェックする(ステップ5c)。そして、一致すればposに次のトークンのポインタをセットし(ステップ5d)、一致しなければステップ5eへ移る。そして、照合結果格納領域(図7)に結果をセットする(ステップ5e)。ここでは、「年」の照合を行ったので、照合結果格納領域の一番左に照合結果をセットする。年パターン・テーブルとトークン列が一致していれば、フラグに1をセットし、「年」表現の先頭と末尾のトークンのオフセットをセットする。以下、同様に、左から「月」「日」「曜日」の照合を行い、結果をセットする(ステップ5f~5g)。

【0032】ここで、「年」「月」「日」の照合時の数値の範囲について幾つか述べる。まず、「年」が和暦表現であった場合、例えば、「昭和」では「1」~「64」というように範囲を設定する必要がある。また、「月」には、「1月」~「12月」しか存在し得ないので、ここにも範囲の設定が必要である。「日」の場合 50 は、月により範囲が異なる。つまり「1・3・5・7・

8・10・12月」は「1」～「31」、「4・6・9・11月」は「1」～「30」、「2月」は、うるう年の場合「1」～「29」、それ以外は「1」～「28」となる。

【0033】日付認定処理を行ったならば、照合結果格納領域(図7)における日付認定結果と日付パターン・テーブル(図6(b))を照合する(ステップ4g)。

【0034】そして、照合結果をチェックし(ステップ4h)、一致していたならば、日付認定結果のトークンのオフセットを用い、日付表現を1つにまとめる(ステップ4i)。一致していなければ、その次のステップ4jへ移る。

【0035】日付表現が認定できたなら日付表現の次のトークンが存在するか否か、日付表現が認定できなかったなら今チェックした次のトークンが存在するか否かをチェックし(ステップ4j)、存在すれば次のステップへ移り、存在しなければ処理を終える。

【0036】次のトークンが存在すれば、日付表現が認定できたなら日付表現の次のトークンのポインタを、日付表現が認定できなかったなら今チェックした次のトークンのポインタをposにセットし(ステップ4k)、ステップ4fへ戻る。

【0037】このようにして、形態素解析での日本語文における日付表現を1つのトークンとしてまとめることができる。

【0038】(実施の形態2)図3は、本発明の日付表現認定及び第2言語への変換に関する実施の形態2における日本語形態素解析装置を示す機能ブロック図である。

【0039】図3に示すように、本実施の形態の日本語形態素解析装置は、入力部(入力手段)31、文分割部(文分割手段)34、トークン認定部(トークン認定手段)35、日付表現認定部(日付表現認定手段)37、文字列変換部(文字列変換手段)41及び出力部42を有している。また、これらを制御する制御部43及び所定情報を記憶する記憶部44が設けられている。さらに、文分割部34には辞書検索部(辞書検索手段)33及び辞書群32が、日付表現認定部37には日付パターン・テーブル36が、文字列変換部41には年号対応テーブル38、月対応テーブル39及び曜日対応テーブル40が、それぞれ設けられている。

【0040】ここで、入力部31、辞書群32、辞書検索部33、文分割部34、トークン認定部35、日付パターン・テーブル36及び日付表現認定部37は、それぞれ図1に対応するものと同様である。また、年号対応テーブル38は、和暦と西暦の対応テーブルである。月対応テーブル39は、第1言語(日本語)と第2言語(第1言語以外の言語)の月に関する対応テーブルである。曜日対応テーブル40は、第1言語と第2言語の曜日に関する対応テーブルである。文字列変換部41は、

日付表現認定部37で認定した日付表現の各パーツ(年号、月、日、曜日)を、年号対応テーブル38、月対応テーブル39、曜日対応テーブル40を用いて第2言語へ変換し、更に、第2言語の日付表現文字列を生成する。出力部42は、トークン認定部35、日付表現認定部38及び文字列変換部41の結果を形態素解析木として出力する。制御部43は、入力部31、文分割部34、トークン認定部35、日付表現認定部37、文字列変換部41、出力部42を制御し、入力部1で入力された文の形態情報を出力する。記憶部44は、入力部31で入力された文、文分割部34における分割結果、トークン認定部35及び日付表現認定部37における認定結果、文字列変換部41における変換結果を記憶する。

【0041】ここで、年号対応テーブル38、月対応テーブル39、曜日対応テーブル40は、プログラム記憶領域であるROMに記憶されている。そして、文字列変換部41、出力部42、制御部43の動作は、CPUがRAMおよびROMとデータのやりとりを行いながら、ROMに記憶されたプログラムを実行することにより実現されている(図2参照)。

【0042】なお、本実施の形態では、CPU23はROM25に記憶されたプログラムを実行する形態となっているが、プログラムはCD-ROMなどの記録媒体27に記録されたものであっても構わない。

【0043】このような形態とした場合は、記録媒体27に記録されているプログラムをディスクドライブ26から読み取り、RAM24に展開したものをCPU23が実行する形態となる。

【0044】このような形態とすることにより、本発明を汎用コンピュータで容易に実現することができる。

【0045】(実施例2)認定した日付表現を第2言語の日付表現に変換する場合も、実施例1と同様の処理の流れで行うことができる。ここでは、実施例1と処理の異なる部分のみを、第2言語を英語として説明する。

【0046】図8は実施の形態2に対応した実施例2における、第1言語を第2言語の表現に変換する対応テーブルを示す説明図、図9は実施例2における第2言語による日付認定を示す説明図である。

【0047】まず、図5の日付を構成する各パーツを認定した際に、対応する第2言語の表現に変換するための対応テーブルを図8のように定義する。ここで、「年」が西暦表現であった場合は、変換の必要はないので、対応テーブルは存在しない。和暦表現であった場合は、年号対応テーブルのベース(右側の値)+認定結果の数値部分-1で西暦の年号が求められる。「月」は、「認定結果の数値部分-1」を添字として、月対応テーブルを参照することで、対応する英語訳が求められる。「曜日」は、認定結果の曜日(「月」～「日」)と曜日対応テーブルの「曜日」を比較し、一致したエントリを参照することで、対応する英語訳が求められる。また、照合

結果格納領域に変換した文字列を格納するためのフィールドを追加する。今、入力文として、

「平成 9 年 10 月 25 日 土曜日」

なる文字列の入力があったとする。これは、トークン認定で、

平成／9／年／10／月／25／日／土曜日

と分割され、図 9 で示すように日付認定される。

【0048】こうして得られた日付表現の各パーツを第 2 言語の表現に並べ替えることで、第 2 言語の日付表現が得られる。

【0049】上記の入力文に対する第 2 言語訳は、Saturday, October 25, 1997 となる。

【0050】このようにして、第 1 言語の日付表現を第 2 言語の日付表現に変換することができる。

【0051】

【発明の効果】以上のように、本発明によれば、形態素解析での日本語文における日付表現を 1 つのトークンとしてまとめることができるという有効な効果が得られる。

【0052】また、本発明によれば、第 1 言語の日付表現を第 2 言語の日付表現に変換することが可能になるという有効な効果が得られる。

【0053】これにより、以降の解析処理の効率化を図ることが可能になるという有効な効果が得られる。

【図面の簡単な説明】

【図 1】本発明の日付表現認定に関する実施の形態 1 における日本語形態素解析装置を示す機能ブロック図

【図 2】図 1 の日本語形態素解析装置の回路ブロック図

【図 3】本発明の日付表現認定及び第 2 言語への変換に

関する実施の形態 2 における日本語形態素解析装置を示す機能ブロック図

【図 4】実施の形態 1 に対応した実施例 1 における、入力文から日付表現を認定する形態素解析処理の流れを示すフローチャート

【図 5】実施例 1 における日付表現認定処理の流れを示すフローチャート

【図 6】実施例 1 における日付表現認定の処理に用いられる日付パターン・テーブルを示す説明図

10 【図 7】実施例 1 における照合結果格納領域を示す説明図

【図 8】実施の形態 2 に対応した実施例 2 における、第 1 言語を第 2 言語の表現に変換する対応テーブルを示す説明図

【図 9】実施例 2 における第 2 言語による日付認定を示す説明図

【符号の説明】

1 入力部（入力手段）

2 辞書群

20 3 辞書検索部（辞書検索手段）

4 文分割部（文分割手段）

5 トークン認定部（トークン認定手段）

7 日付表現認定部（日付表現認定手段）

31 入力部（入力手段）

32 辞書群

33 辞書検索部（辞書検索手段）

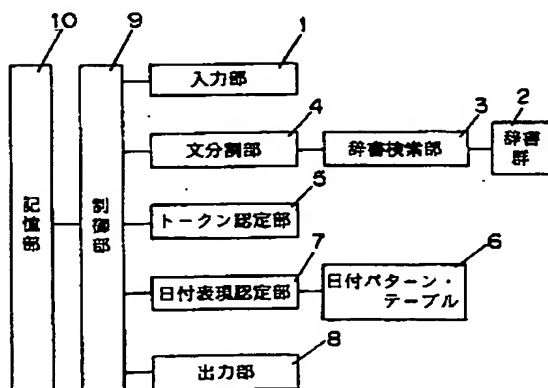
34 文分割部（文分割手段）

35 トークン認定部（トークン認定手段）

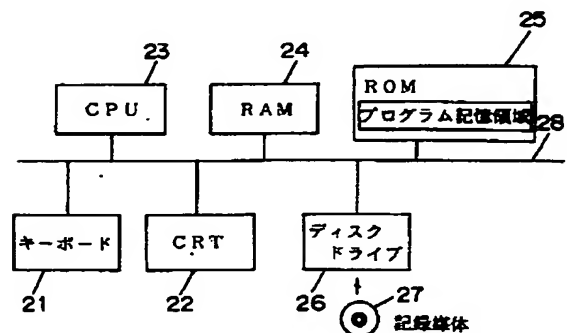
37 日付表現認定部（日付表現認定手段）

30 41 文字列変換部（文字列変換手段）

【図 1】



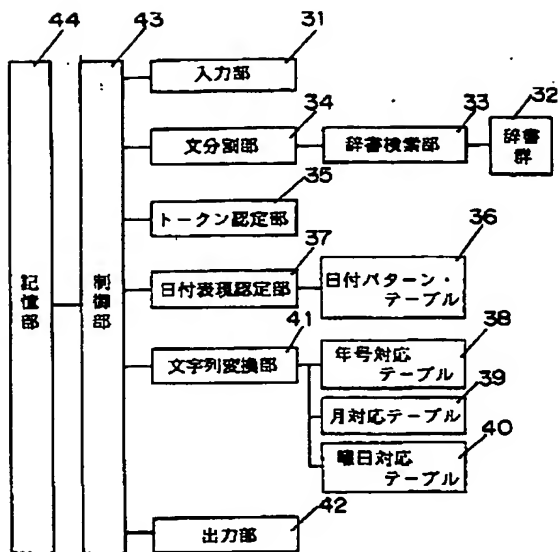
【図 2】



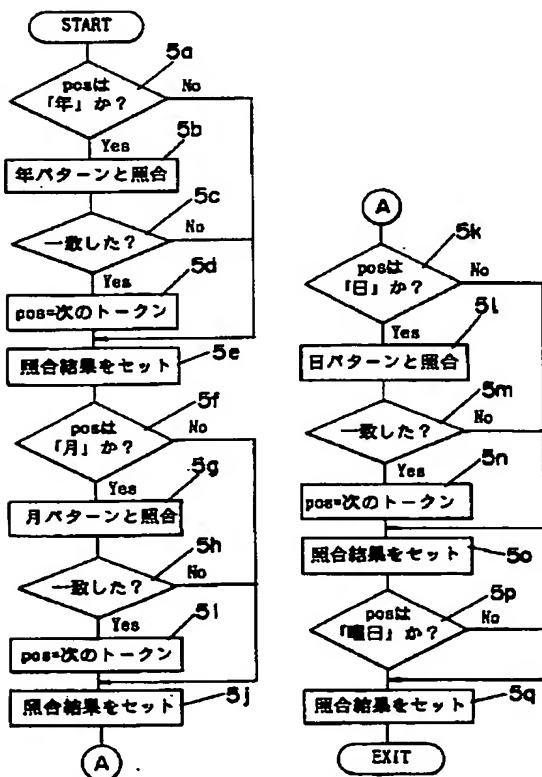
【図 9】

年	月	日	曜日
1	1	1	1
0 2	3 4	5 6	7 7
1997	October	25	Saturday

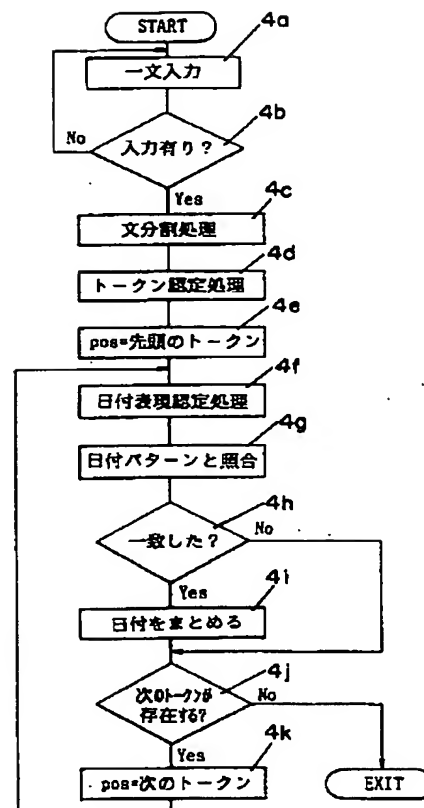
【図3】



【図5】



【図4】



【図6】

(a)

年パターン・テーブル:

元号	数	「年」
元号	「元」	「年」
数	「年」	

月パターン・テーブル:

数	「月」
---	-----

日パターン・テーブル:

数	「日」
---	-----

曜日パターン・テーブル:

曜日		
(曜日)

(b)

年	月	日	曜日
1	1	1	1
1	1	1	0
1	1	0	0
0	1	1	1
0	1	1	0
0	0	1	1

【図 7】

年		月		日		曜日
1		1		1		0
3	4	5	6	7	8	

「年」表現の末尾のトークン

「年」表現の先頭のトークン

【図 8】

年号対応テーブル:

平成	1989
昭和	1926
大正	1912
明治	1868

曜日対応テーブル:

月	Monday
火	Tuesday
水	Wednesday
木	Thursday
金	Friday
土	Saturday
日	Sunday

月対応テーブル:

January
February
March
April
May
June
July
August
September
October
November
December